

# Trade-off between disease resistance and crop yield: a landscape-scale mathematical modelling perspective

September 5, 2016

## Abstract

The deployment of crop varieties that are partially resistant to plant pathogens is an important method of disease control. However, a trade-off may occur between the benefits of planting the resistant variety and a yield penalty, whereby the standard susceptible variety out-yields the resistant one in the absence of disease. This presents a dilemma: deploying the resistant variety is advisable only if the disease occurs and is sufficient for the resistant variety to out-yield the infected standard variety. Additionally, planting the resistant variety carries with it a further advantage in that the resistant variety reduces the probability of disease invading. Therefore, viewed from the perspective of a grower community, there is likely to be an optimal trade-off and thus an optimal cropping density for the resistant variety. We introduce a simple stochastic, epidemiological model to investigate the trade-off and the consequences for crop yield. Focusing on SIR epidemic dynamics, we use the final size equation to calculate the surviving host population in order to analyse the yield, an approach suitable for rapid epidemics in agricultural crops. We identify a single compound parameter which we call the efficacy of resistance and which incorporates the changes in susceptibility, infectivity and durability of the resistant variety. We use the compound parameter to inform policy plots that can be used to identify the optimal strategy for given parameter values when an outbreak is certain. When the outbreak is uncertain, we show that for some parameter values planting the resistant variety is optimal even when it would not be during the outbreak. This is because the resistant variety reduces the probability of an outbreak occurring.

# 1 Introduction

Plant diseases impose significant crop losses in agriculture, horticulture and forestry. Estimates of annual crop losses in agriculture range from 14% of crop yield worldwide [1–3], extending to 20–40% when weeds and pests are included [4], with total elimination of crops in some severe epidemics [5]. The deployment of crop varieties that are genetically resistant to plant pathogens provides an efficient means of disease control. The simplest case involves complete resistance to a pathogen that remains durable for long periods of time without being overcome by new virulent strains of the pathogen [6, 7]. Newly released resistant varieties should ideally also match or surpass the agronomic properties of the susceptible varieties that they replace. Historically many resistant varieties exhibited a qualitative resistance that is a form of complete resistance in which the pathogen is unable to infect the host [8]. There is increasing awareness, however, that qualitative resistance under single gene control in the host imposes such strong pressures on the pathogen population to overcome resistance so that failure becomes almost inevitable [9, 10]. Accordingly, increasing attention is being paid to the release of partially resistant varieties that slow or otherwise reduce, but do not prevent, infection and multiplication of the pathogen on the host crop [10, 11]. Such resistance is frequently under the control of few to many genes in the host crop [12, 13]. Trade-offs may occur, however, between partial resistance and the agronomic properties of the host crop such that partial resistance may be associated with lower yield for reduced quality in the crop variety compared with the standard [14–17]. This presents individual growers with a dilemma: growing the resistant crop imposes a yield penalty that may be compensated for only if disease occurs and is sufficient for the infected, partially resistant crop to out-yield an equivalently infected standard crop. The decision is further complicated: as more growers elect to plant the partially resistant crop the probability of a severe epidemic occurring decreases. Viewed from the perspective of a community of growers there is likely to be an optimal trade-off and hence a cropping density for the proportion of sites that are planted to the partially resistant variety. Moreover, even if an outbreak is certain, it is not necessary for all growers to have planted the partially resistant variety for the benefit to be realised by the community of growers as a whole [18, 19].

Previous epidemiological analyses on the introduction of partially resistant varieties in the landscape have focused on the effects of disease dynamics [18–21]. Here we focus on the trade-off between disease and yield at the landscape scale. The primary aim of this paper is, therefore, to provide insight into when it is worth deploying a partially resistant variety and in what proportions relative to a standard, higher yielding, susceptible variety. We do this using simple stochastic epidemiological models that characterise the spread of the pathogen, and hence disease, and the consequences for crop yield. We identify policy plots [22] that can be used to infer optimal strategies given some prior knowledge of the resistance-yield trade-offs. We make certain simplifying assumptions that we subsequently relax. Our intention in this paper is to address the generic problem of how to deploy partial resistance when there is a yield penalty and uncertainty. We intentionally propose a flexible modelling framework that encompasses a variety of epidemiological mechanisms that could be associated with partial resistance, with broad applicability to a range of host pathogen systems. Hence, we consider SIR epidemics, in which susceptible hosts ( $S$ ) become infected ( $I$ ), remain infectious for a period of time and then are removed ( $R$ ). Removal may occur naturally by disease-induced death or by deliberate removal, for example by roguing of infected plants. SIR epidemics are typified by an increase in infected hosts followed by a decrease as the epidemic ‘burns itself out’ or is controlled. We assume that yield is a function of the amount of healthy, i.e. uninfected host, allowing for a yield penalty for healthy hosts of the partially-resistant compared with the susceptible variety. We initially assume that yield is accumulated over a long period of time relative to the period of crop growth and the time course of the epidemic. This assumption, enables us to gain analytical insight to inform the deployment of partially-resistant crops subject to yield penalties. We subsequently relax the assumptions to test the robustness of the conclusions. Our initial results apply to annual crops, in which epidemics happen fast, typified by potato late blight and rusts of small grain cereals. Here,  $S$  and consequently  $I$ , are expressed as units of plant host tissue [23]. Our results also apply to cassava virus diseases, in which yield is accumulated over a long period and roguing is practiced to remove infected hosts. In this case,  $S$  and  $I$  refer to whole plants. The model may also be applied, more generally, to perennial crops in which roguing of infected plants occurs, yield accumulates over long periods and there is continuing harvesting. We first consider the case where epidemics are inevitable and address

two broad questions: under what circumstances is investing in the partially resistant variety likely to be profitable, and how much of the resistant variety should be deployed to maximise yield? We subsequently consider the robustness of the inferences about the deployment of the partially resistant variety when the occurrence of the epidemic is uncertain.

## 2 Methods

### 2.1 Epidemic model

We consider disease spreading through a metapopulation, comprising two types of host crop, a partially resistant and a fully susceptible 'standard' variety. Plants of each variety can be in one of three classes: susceptible (S, i.e. healthy), infected (I) or removed (R, i.e. post-infectious). The principal parameters used in the models for disease spread and yield are summarised in Table 1, for ease of reference. We make the following assumptions:

- The epidemic follows *SIR* compartmental dynamics, with density-dependent mixing [24]. The  $I \rightarrow R$  transition is realized either by disease-induced mortality, by roguing of the infected hosts, or by a combination of both.
- Hosts in the infected and removed classes do not contribute to the yield.
- The resistant hosts are less likely to become infected upon contact with the pathogen (by a factor  $\eta \in [0, 1]$ ), produce less inoculum (by a factor  $\nu \in [0, 1]$ ), have a different infectious period (by a factor  $\sigma \in (0, \infty)$ ) and contribute less to the final yield (by a factor  $f \in [0, 1]$ ). It is natural to consider both shorter and longer infectious periods for the resistant variety, because the resistant hosts might take longer to die or they might take longer to develop visible symptoms upon infection and thus avoid detection and subsequent removal (by, for example, roguing of symptomatic hosts in certain crops) for longer than the standard hosts.

| Parameter, Value/Range      | Description  |
|-----------------------------|--|
| $\rho, [0, 1]$              | Proportion of the resistant hosts in the population.                                     |
| $\beta$ , both 2 and 3 used | Transmission rate between the standard hosts.  |
| $\mu, 1$                    | Infectious period of the standard hosts set to 1.  |
| $\eta, [0, 1]$              | <i>Susceptibility factor</i> : reduction of susceptibility of the resistant hosts.       |
| $\nu, [0, 1]$               | <i>Infectivity factor</i> : reduction of infectiousness of the resistant hosts.          |
| $f, [0, 1]$                 | <i>Yield penalty</i> : reduction of yield of the resistant hosts.                        |
| $\sigma, (0, \infty)$       | <i>Removal factor</i> : change in the removal rate of the resistant hosts.               |
| $R_0$ , both 2 and 3 used   | Basic reproductive number in a population of standard hosts.                             |
| $\xi, [0, 1]$               | <i>Resistance efficacy</i> , $\xi = 1 - \eta\sigma\nu$ , see the <i>Results</i> section. |
| $\lambda, 2$                | Probabilistic rate of import of the pathogen, see Section 2.3.                           |

Table 1: *List of all the parameters used.*

We assume a fraction  $\rho = N_R/(N_S + N_R)$  of all the hosts to be resistant, where  $N_S$  and  $N_R$  are the numbers of standard and resistant hosts, respectively. The model for disease spread is given by (where subscripts S and R denote standard and resistant varieties):

$$\dot{S}_S = -\beta S_S[(1 - \rho)I_S + \nu\rho I_R] \quad (1)$$

$$\dot{S}_R = -\eta\beta S_R[(1 - \rho)I_S + \nu\rho I_R] \quad (2)$$

$$\dot{I}_S = \beta S_S[(1 - \rho)I_S + \nu\rho I_R] - \mu I_S \quad (3)$$

$$\dot{I}_R = \eta\beta S_R[(1 - \rho)I_S + \nu\rho I_R] - \frac{\mu}{\sigma} I_R \quad (4)$$

$$S_S + I_S + R_S = S_R + I_R + R_R = 1, \quad (5)$$

where  $\beta$  is the transmission rate and  $\mu$  ( $\mu/\sigma$ ) is the removal rate of the standard (resistant) hosts respectively. It is convenient to introduce the basic reproductive number of the standard hosts in a

monoculture,  $R_0 = \beta/\mu$ . This is defined in the standard way as the average number of secondary infections caused by a single infected host over the course of its infectious period.

## 2.2 Yield without uncertainty

When an outbreak is certain, we represent the yield *per host* and *per unit time*, accumulated by healthy plants over a fixed period  $T$ , by a straightforward adaptation of the integral over the susceptible hosts, a simple measure used in plant epidemiology [25, 26]

$$Y = \frac{1}{T} \int_0^T [(1 - \rho)S_S(t) + f\rho S_R(t)]dt. \quad (6)$$

To simplify the analysis we consider the yield accumulated over a very long period of time period relative to the duration of the longest epidemic ( $T \rightarrow \infty$  in the above equation), so that it is effectively given by the proportions of susceptible hosts that survive the epidemic. We include an explanatory sketch in the *Supplementary Materials* S1. Therefore the yield  $Y$  is given by

$$Y = (1 - \rho)S_S^\infty + f\rho S_R^\infty. \quad (7)$$

where  $S_i^\infty$  is the proportion of the susceptible hosts in the population  $i \in \{S, R\}$  after the outbreak has ended. Note that this assumption is made purely for mathematical convenience to make analytical progress, from which initial insights into the optimal strategy can be inferred. We subsequently relax this assumption. We introduce an arbitrary, finite-time horizon,  $T_E$  to Eqn 6 and show that the inferences derived for an infinite-time horizon hold when yield is accumulated over a finite period of time (Section 3.4). The choice of  $T_E$ , while arbitrary, is motivated by keeping the analysis generic.

## 2.3 Yield with uncertainty

When there is uncertainty about whether or not an outbreak will occur, the yield function is given by the expected yield:

$$Y(\rho) = p_E \times \text{Yield}(\text{outbreak}) + (1 - p_E) \times \text{Yield}(\text{no outbreak}) \quad (8)$$

$$= p_E[(1 - \rho)S_S^\infty + f\rho S_R^\infty] + (1 - p_E)[1 - \rho + f\rho], \quad (9)$$

where  $p_E$  is the probability of an epidemic. To model this probability, we consider a situation in which the pathogen has a constant small rate of introduction into the host population over some period of time (for example because the climatic conditions are favourable during this period). The number of introductions then follows a Poisson distribution with mean  $\lambda_1$ . Each time the pathogen is introduced into the system, it infects an initial host with probability  $\lambda_2(1 - \rho + \eta\rho)$  where  $\lambda_2$  is the probability of a standard host getting infected upon contact with the pathogen. For convenience, we define  $\lambda = \lambda_1\lambda_2$ . Once the initial host has been infected, a large-scale outbreak will occur with probability  $P_{\text{takeoff}}$  which can be calculated using standard arguments, see *Supplementary Materials* S2. Putting this together and using the thinning property of Poisson processes leads to the expression for the overall probability of an epidemic as

$$p_E = 1 - e^{-\lambda(1 - \rho + \eta\rho)P_{\text{takeoff}}}. \quad (10)$$

## 2.4 Summary of the assumptions

In this section, for convenience and reference, we summarise the principal assumptions of the model and the subsequent analyses. We approach the problem generically, using biologically plausible parameter values and ranges to reflect classes of host-pathogen system rather than restricting the analysis to a single system. Instead, our analysis is designed to identify which parameters are important and what are their critical ranges. We introduce flexibility by allowing for the epidemiological mechanisms accounting for partial resistance to be expressed through changes in one or more of the following: the infectivity of infected hosts, the susceptibility of healthy hosts or the length of the infectious period. These effects can be tuned independently in our model through different parameters. In our model the unit of host can be either a whole plant or healthy tissue and we assume the infected hosts eventually die or are

removed through roguing. For simplicity we assume that neither infected nor removed hosts contribute to the yield. This means there is no replanting or that it takes a long time for a replanted host to reach maturity, such as in the case of tree crops. The yield is modelled as an integral over the healthy hosts. Initially, in order to make the model analytically tractable we assume the yield is accumulated over a time-scale much longer than the epidemic duration so that it can be approximated by an infinite time horizon. In the second part of the analysis, we relax this assumption and consider yield accumulated only over the duration of the epidemic. The main purpose of this is to verify that our qualitative results are not simply an artefact of the infinite time horizon. Finally, when we allow for uncertainty in the occurrence of an epidemic, we assume that the pathogen has a constant rate of import into the host population over a certain time period preceding the potential epidemic.

### 3 Results

#### 3.1 Model without uncertainty

Using the final size equations for an epidemic [24], it is possible to derive an analytic expression for the yield function (see *Supplementary Materials S3* for details),

$$Y(x) = x - (x - fx^\eta) \frac{\ln(x) + R_0(1-x)}{R_0[1-x - \nu\sigma(1-x^\eta)]}, \quad (11)$$

where  $x \equiv S_S^\infty$  is the final size of the susceptible class of the non-resistant hosts and depends on  $\rho$ . It is not possible to obtain an analytic expression for the yield as a function of  $\rho$  directly. However, the analytic solution (i.e. 11) shows three possible control scenarios to optimise yield. These are illustrated for three different yield penalties associated with the partially resistant variety in Fig. 1. There are two extreme scenarios: no control, i.e. grow only the standard variety when the yield penalty is high (Fig. 1A) and 'full' control, i.e. grow sufficient resistant variety to bring the basic reproductive number below 1 (see below) when the yield penalty is low (Fig. 1C). As the yield penalty increases, so the cropping ratio of standard to resistant variety required to achieve an optimal yield increases leading to an intermediate control scenario (Fig. 1B). Note that the 'full' control (Fig. 1C) does not necessarily mean  $\rho = 1$  but

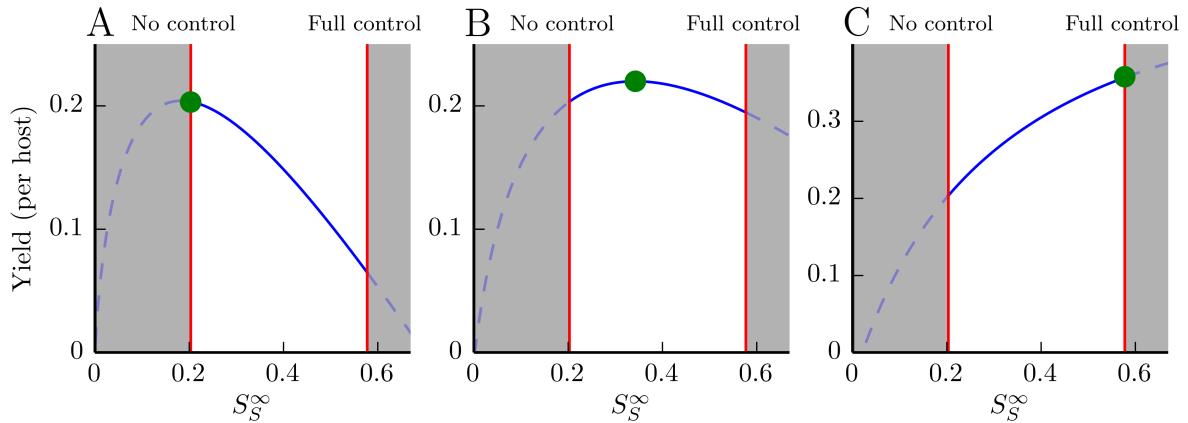


Figure 1: *Three possible control scenarios, and optimal yields illustrated for three yield penalties associated with a partially resistant variety,  $f=0.1$  (A),  $f=0.3$  (B) and  $f=0.55$  (C). The vertical red lines show the range of possible results, that is they show the final sizes corresponding to no control and to full control. The green dot shows the optimal final size in each case. Default parameter values:  $R_0 = 2$ ,  $\eta = 0.79$ ,  $\nu\sigma = 0.78$ . The values were selected in order to illustrate the three distinct types of behaviour.*

corresponds to the density  $\rho$  being equal to

$$\rho = \rho_{\text{full}} = \min(1, \rho_c), \quad (12)$$

where  $\rho_c$  is a critical density such that the effective basic reproductive number of the system (1-4)  $R'_0$  in the presence of the resistant hosts falls below 1. This is because once the basic reproductive number falls below one, the epidemic is prevented and further deployment of the resistant variety has no effect. The effective basic reproductive number  $R'_0$  can be calculated using the next generation method [27]: it is given by

$$R'_0(\rho) = R_0(1 - \rho\xi), \quad (13)$$

where we introduce the parameter  $\xi = 1 - \eta\sigma\nu$  which we will refer to as the *resistance efficacy*.

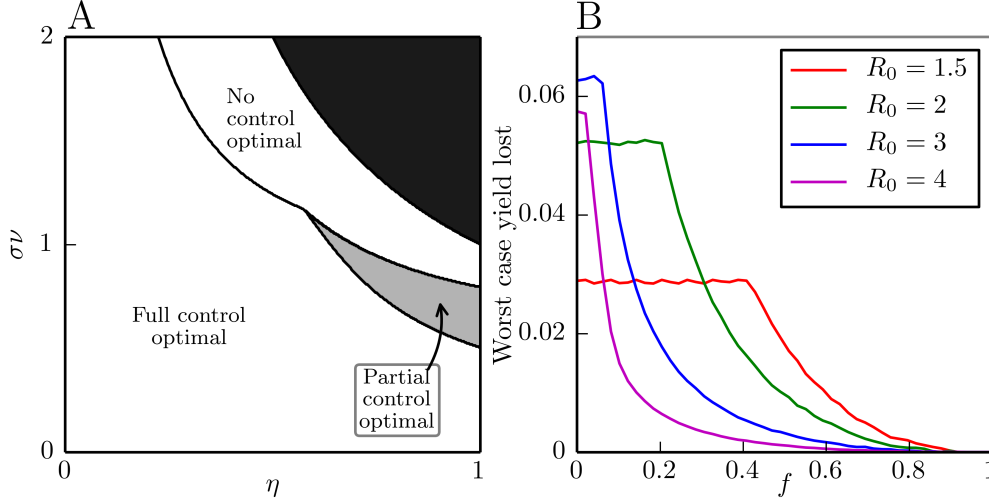


Figure 2: **A** Policy plot for  $f = 0.3$  and  $R_0 = 3$ . The dark region corresponds to  $\xi < 0$  that is controls that increase the basic reproductive number and support the spread of the pathogen, which we do not consider further. **B** shows how much yield per host is lost when we ignore the intermediate control, in the worst case scenario. Mathematically, the function plotted is  $\max_{\eta, \sigma\nu}(Y_{\text{optimal}} - \max(Y_{\text{full}}, Y_{\text{none}}))$ .

From equation (11) we conclude that the optimal proportion  $\rho$  only depends on  $\nu$  and  $\sigma$  through their product  $\nu\sigma$ . This allows us to make 2D policy plots showing where different types of control are optimal, for various values of yield penalty,  $f$  (see Fig. 2 for an example). Note that when  $\xi = 1 - \eta\sigma\nu < 0$  (the dark region in Fig. 2A), the control supports the spread of the pathogen and we therefore do not consider this region any further. Examination of Fig. 2A and corresponding figures for a range of values of  $f$  shows that the region of the parameter space where intermediate control is optimal is small. Accordingly we investigate the potential for loss in yield if the option of intermediate control is ignored, using  $Y_{\text{optimal}} - \max(Y_{\text{full}}, Y_{\text{none}})$  as a metric for yield loss in choosing no or 'full' control in place of intermediate control (*cf.* Fig. 1). The metric is a function of parameters  $f$ ,  $\eta$ ,  $\nu$  and  $\sigma$ . In Figure 2B, we plot the worst yield loss as a function of  $f$ , that is  $\max_{\eta, \sigma\nu}(Y_{\text{optimal}} - \max(Y_{\text{full}}, Y_{\text{none}}))$ . The losses are small and decrease very quickly with  $f$ . Furthermore, in the  $(\eta, \nu\sigma)$  plane they are only appreciable along the curve  $Y_{\text{full}} = Y_{\text{none}}$  and close to  $\eta = 1$ , where the relative susceptibilities of resistant and standard varieties are identical. The explanation of this can be found in *Supplementary Materials S4*. This means that for the most part, we can focus on the extreme controls since they are mostly optimal and when they are not, they provide a good approximation to the optimum.

### 3.2 Extreme control optimization

From equation (11) we can derive the conditions for deployment of a resistant variety under which 'full' control is better than no control. There are two cases to consider. If the resistance is effective enough so that the outbreak can be prevented altogether, that is if  $R'_0(\rho = 1) = (1 - \xi)R_0 < 1$ , full control is better than no control when

$$\xi > \frac{(1 - 1/R_0)(1 - f)}{1 - s_0} \quad (14)$$

where  $s_0$  is the proportion of surviving hosts when no control is deployed. If the resistance is not effective enough to prevent the outbreak, that is if  $(1 - \xi)R_0 > 1$ , full control is better than no control when

$$\xi > 1 - \frac{f}{R_0} \frac{\ln(f/s_0)}{f - s_0}, \quad (15)$$

These conditions are derived in the *Supplementary Materials* S4. Note that these conditions only depend on the parameters  $\eta$ ,  $\sigma$  and  $\nu$  via the resistance efficacy  $\xi = 1 - \eta\sigma\nu$  and therefore there are effectively only 3 controlling parameters:  $R_0$ ,  $\xi$  and  $f$ . This allows us to plot a policy diagram, showing when the resistant variety is worth deploying. The plot for  $R_0 = 3$  can be found in Fig. 3; other basic reproductive numbers are not qualitatively different. Figure 3A shows the region where full control is optimal and the boundary, which corresponds to the critical value of  $\xi$ .

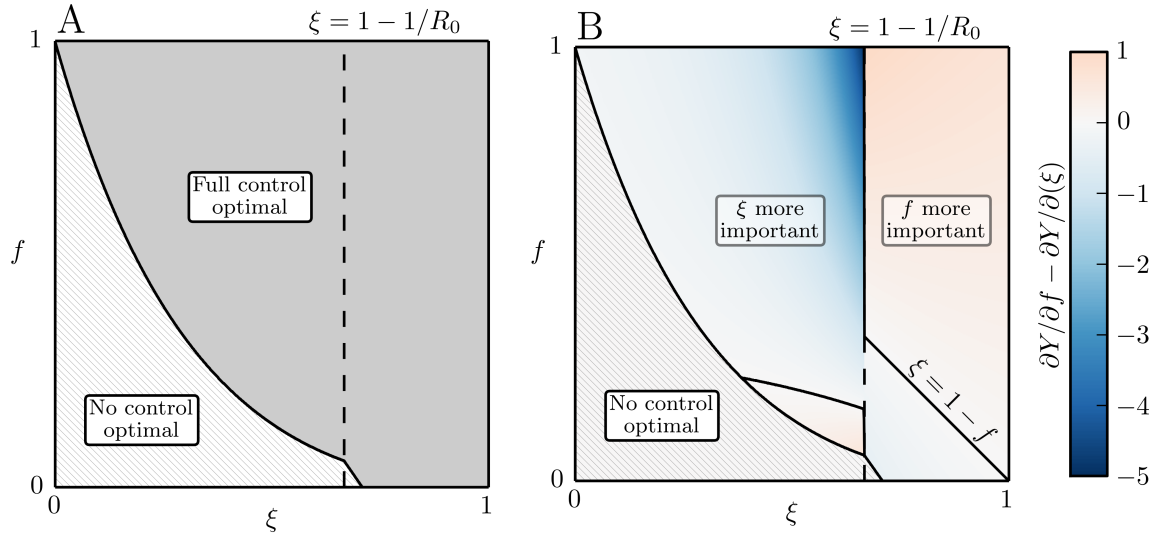


Figure 3: **A** Policy plot for  $R_0 = 3$  to inform control scenarios when the resistance efficacy and yield penalty for the resistant variety are known. **B** Policy plot to show sensitivity of yield at optimal strategy to changing  $f$  (yield penalty) and  $\xi$  (resistance efficacy). The colour scale corresponds to the difference of these two rates,  $\frac{\partial Y}{\partial f} - \frac{\partial Y}{\partial \xi}$ . The additional black lines in 3B mark the boundaries of the blue and the red regions.

Figure 3B shows whether the optimal value of the yield,  $Y$ , increases more quickly as the yield penalty  $f$  increases or when the resistance efficacy,  $\xi$ , increases, using the quantity  $\frac{\partial Y}{\partial f} - \frac{\partial Y}{\partial \xi}$  as a metric. When the metric is positive, increasing  $f$  is more important and when it is negative increasing  $\xi$  is more important to increase yield. The effect of changing  $\xi$  is most dramatic near the boundary  $\xi = 1 - 1/R_0$  and therefore when  $\xi$  is close to this boundary from below, it is much more profitable to increase it above the boundary than to attempt to increase the yield factor.

### 3.3 Model with uncertainty

The probability of invasion,  $p_E$ , derived from equation (10) decreases as the cropping ratio of the resistant hosts increases (Fig. 4). The decline in  $p_E$  is steeper when the susceptibility factor ( $\eta$ , Table 1) of the resistant fraction of the host population becomes smaller, that is, when this fraction of hosts presents an increasing level of resistance. To investigate the impact on the yield of the deployment of the resistant variety in the presence of uncertainty, we repeat our analysis from the previous section. Similarly to the case of a deterministic, certain outbreak, numerical analysis reveals that a good approximation is provided by only considering extreme controls, that is none or 'full'. Note that because of the introduction of the probability  $p_E$  there are now 4 independent controlling parameters,  $\xi$ ,  $R_0$ ,  $f$  and  $\eta$ . In Figure 5 we show the policy plots corresponding to three different values of  $\eta$  and  $R_0 = 3$  for both the deterministic and the stochastic case. We can see that when the resis-

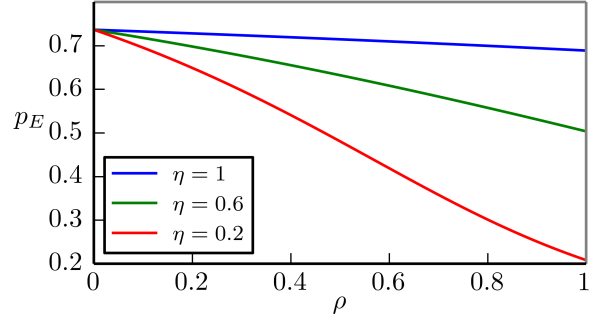


Figure 4: *Probability of an outbreak as a function of the proportion of the resistant hosts in the population. Parameters are  $\lambda = 2$ ,  $\xi = 0.2$  and  $R_0 = 3$ .*

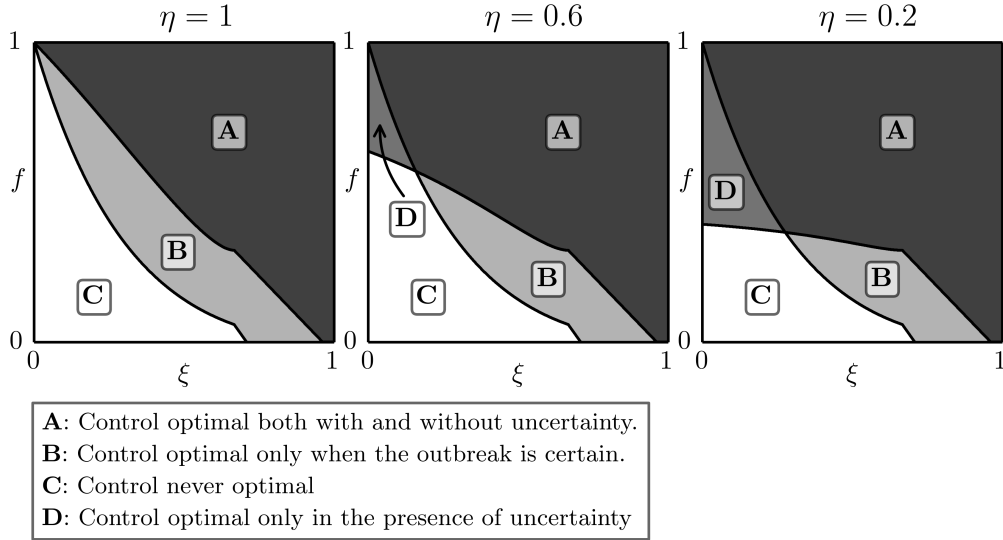


Figure 5:  $\lambda = 2$ ,  $R_0 = 3$ . *Policy plots for three different values of  $\eta$ , with and without allowance for uncertainty in an outbreak. Note that when the resistance is relatively ineffective ( $\xi$  low), but the resistant hosts are significantly less susceptible to the pathogen, full control becomes optimal when an outbreak is uncertain when it would not be optimal in the deterministic setting.*

tance significantly reduces the susceptibility to the pathogen, but its efficacy  $\xi$  is low overall because  $\sigma > 1$ , it can happen that the control is optimal only in the presence of uncertainty. This means that while the control is not desirable during the outbreak, the benefits of the possibility of preventing the outbreak altogether are significant. We can formalize this by defining  $\Delta Y = Y(\text{Optimal}) - Y(\text{No control})$ , that is the yield gained by controlling correctly as opposed to not controlling at all. In Figure 6, we plot the difference between the yield gained by controlling with uncertainty and without uncertainty. We can see that without uncertainty the control provides greater yield gains for high resistance efficacy but, in agreement with Figure 5, when the resistance efficacy is low, the susceptibility factor  $\eta$  is low and the yield parameter  $f$  is high, the control provides significantly greater yield gains when the uncertainty is



present. This demonstrates the importance of reducing the probability of an outbreak occurring even when during the outbreak the control is not desirable.

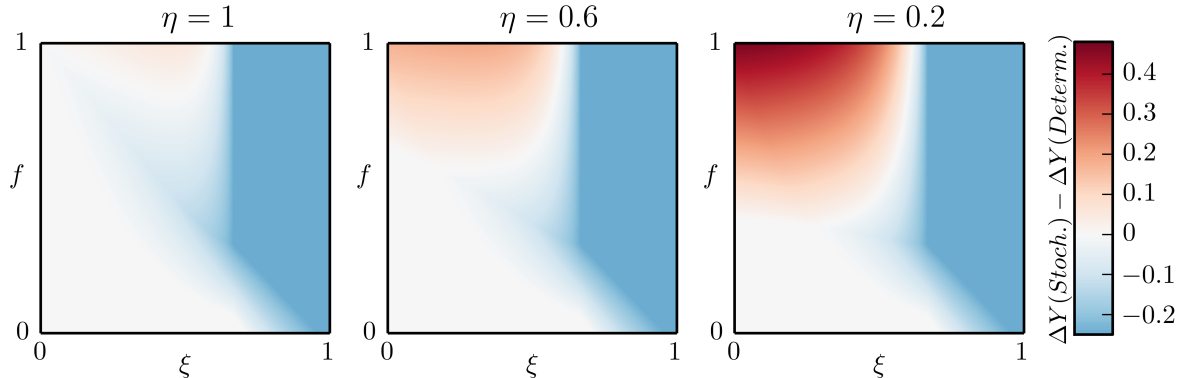


Figure 6:  $\lambda = 2$ ,  $R_0 = 3$ . The colour shows how much more yield is gained by the control when the uncertainty is present over when the outbreak is certain, mathematically  $\Delta Y(\text{stochastic}) - \Delta Y(\text{deterministic})$ . We can see that when the resistance efficacy is low,  $f$  is high and the susceptibility is significantly reduced by the resistance, the control provides higher gains in yield when the uncertainty is present. This demonstrates the importance of reduction of the probability of an outbreak occurring even when during the outbreak the control is not desirable.

### 3.4 Yield model with finite time horizon

So far we have considered long-term yield. We have shown that contrary to intuition, the benefit of deploying a resistant variety can be greater when the disease outbreak is uncertain. We now verify this result when the assumption of the infinite time horizon is relaxed and the yield to be maximized is evaluated over a finite period of time. We selected the time of the duration of the epidemic  $T_E$  as a representative of this finite time horizon. Therefore the yield is now given by

$$Y = \frac{1}{T_E} \int_0^{T_E} [(1 - \rho)S_S(t) + f\rho S_R(t)]dt. \quad (16)$$

Whereas our initial model focuses on the final state of the epidemic, the beginning of the epidemic and how quickly the pathogen invades now become important factors. Note that since increasing the cropping ratio  $\rho$  of the resistant variety leads to a decrease in  $R_0$  of the system, the duration of the epidemic  $T_E$  depends on  $\rho$ . It is no longer possible to use the final size of the epidemic to characterize the yield in this case and no analytic expression for the yield can be derived. To proceed, we ran numerical simulations of the model for randomly generated parameter values (see the *Supplementary Materials* S5 for the technical details). The results (Fig. 7) are in good agreement with the idealized 'long-time' yield model (cf Fig. 6). As expected, the benefits of full control when an outbreak is uncertain correlate with lower values of  $\eta$  and higher values of  $\sigma$ . The average value of  $\eta$  in the simulations where the uncertainty leads to greater benefits of control was 0.48 while in those where the benefits of control were greater without uncertainty it was 0.63. For  $\sigma$  the values were 4.47 and 1.65 respectively.

## 4 Discussion

We have analysed the impacts on disease dynamics and crop yield at landscape scales of the resistance-yield trade-off when attempting to control a disease by deploying a resistant crop variety in two different situations. In both cases, we assume that the resistant variety carries a yield penalty, i.e. it yields less than the standard variety in the absence of disease. First, we assumed that an epidemic is inevitable in order to analyse how the deployment of the resistant variety impacts on disease dynamics and how these

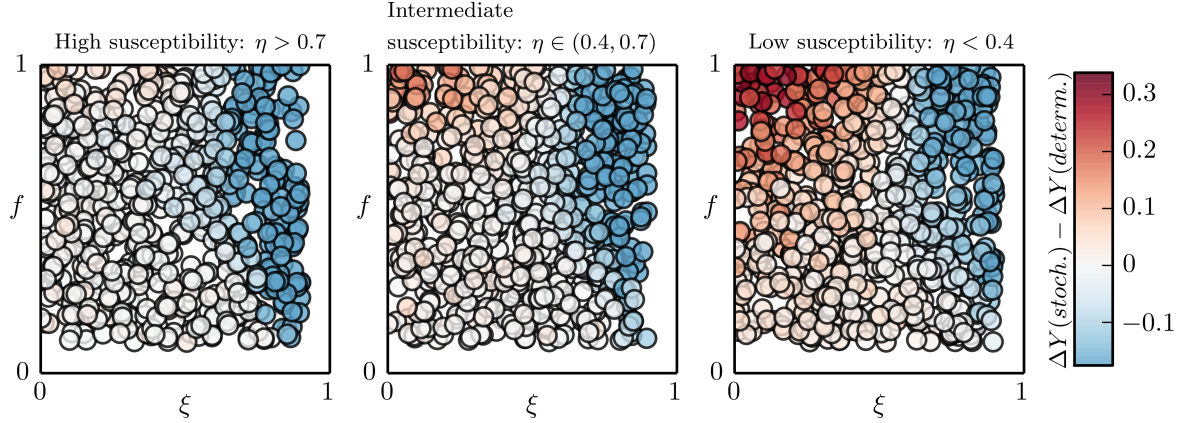


Figure 7:  $R_0 = 3$  and  $\lambda = 2$ . 2000 simulations with random parameter values out of which 1063 lead to the benefit of control being greater with the uncertainty present. The circles show the simulation points that lie in the appropriate interval for  $\eta$ . The colour shows the magnitude of the effect, that is how much greater is the benefit of the control with the uncertainty present. We can see that in agreement with the Figure 6, the effect is the strongest in the upper left corner of the  $\xi$ - $f$  plane and for low values of  $\eta$ .

in turn affect the final yield averaged over multiple fields in the landscape. Subsequently, we considered the situation where only the probability of an epidemic is known at the time of planting. We have not considered detailed models for demographic or environmental stochasticity [24] other than by allowing for uncertainty as to whether an outbreak will occur, coupled with deterministic dynamics. We selected this approach for two reasons. Firstly, it concentrates on the main component of variability that we wanted to investigate, that is whether or not an epidemic occurs. Secondly, it allows us to carry over most of the methodology from the analysis of the problem without uncertainty and it simplifies the numerical analysis when a different yield model with a finite time horizon is considered. We used a simple approach of integrating over the susceptible hosts or host tissue as appropriate [25, 26, 28], which assumes that healthy tissue contributes to yield. Our intention throughout is to introduce a generic modelling framework. The framework is motivated by fast epidemics on agricultural crops in which the epidemic naturally burns itself out or by removal of infected hosts by roguing as may occur, for example, in cassava crops. The framework can, in principle, be extended to perennial crops. To simplify the mathematics and to minimize the number of free parameters, we initially assume that the yield is accumulated over a time period which is much longer than the time-scale of the epidemic such as in the case of potato late blight and some rust diseases of small grain cereals. This allows us to approximate the yield by the amount of susceptible host that survives the epidemic.

When an outbreak is certain, we showed that the susceptibility, infectivity and the removal factors  $\eta$ ,  $\nu$  and  $\sigma$  introduced in the model can be combined into a single parameter  $\xi = 1 - \eta\sigma\nu$  which we call the *efficacy of the resistance*. This aggregate parameter is a convenient means of integrating the components that characterise the differences between resistant and susceptible varieties. The approach is analogous to, yet different from, the approach advocated by Parlevliet [29] directed at quantifying the components of resistance when comparing different varieties. Thus, Parlevliet [29] first showed how to quantify epidemiological components of resistance such as infection frequency, latent period and spore production per unit time as well as the infectious period. From these analyses, Parlevliet [29] was able to quantify and ascribe the components of resistance that accounted for differences between susceptible and resistant hosts. Savary *et al.* [30] subsequently showed for peanut rust how to combine individual components into a product as a relative measure of resistance that reflected differences between epidemiological components, which is analogous to  $\xi$ . The difference lies in that  $\xi$  is constructed so that the components relate directly to parameters that define rates in a epidemiological dynamical model. Thus  $\nu$  is a measure of the reduction in the rate of transmission of infection;  $\sigma$  is a measure of the change in

the infectious period and  $\eta$  is the measure of reduction in the transmission rate. We found the conditions that  $\xi$  has to satisfy in order for deployment of the lower yielding, resistant variety to be optimal. Major gains can be achieved particularly when  $\xi$  increases above  $1 - 1/R_0$ , which is also the condition for being able to prevent the outbreak altogether.

A further trade-off arises when uncertainty about an outbreak is added. Deployment of the resistant variety is wasteful in the absence of an epidemic. However, given uncertainty about an epidemic outbreak, deploying a resistant variety renders the overall host population less susceptible to the pathogen and therefore decreases the probability that an outbreak will occur in the first place [31]. Fig. 5 and Fig. 6 provide the resolution to this trade-off.

Our results indicate that when the resistance is sufficient to prevent an epidemic altogether ( $\xi$  large), the benefits of the control are greater when an outbreak is certain. However, when the resistance is not strong enough to prevent the outbreak ( $\xi$  small) but offers significant reduction in susceptibility to the pathogen ( $\eta$  is small), the benefits of control are overall greater when uncertainty is accounted for (cf Fig. 6 with Fig. 5). In such a case, reducing the invasion probability and thus possibly preventing the outbreak altogether outweighs the risk of wasting resources by deploying resistant cultivars. Note that this is independent of the assumption that the invasions follow a Poisson process. Rather it is a consequence of the fact that when the resistance is not very effective during the outbreak, it can still significantly reduce the probability of the pathogen invading. Biologically, this can happen when the resistant hosts have a longer infectious period (that is  $\sigma > 1$ ). This is possible, for example, when the disease eventually kills the hosts, but takes longer to kill the resistant hosts. Alternatively, in the analogous case when the disease does not kill the hosts but rather they are removed from the population via roguing, the resistant hosts might take longer to show symptoms and therefore avoid detection.

We have made a number of important simplifying assumptions about the epidemiological model. These have allowed some insights to be gained about the trade-offs to be considered in deploying resistant varieties with lower yield potential compared with a susceptible variety. We have used a simple SIR epidemiological model (Eqns 1-5) that is parsimonious while allowing flexibility in attributing the effects of partial resistance to different epidemiological processes. Accordingly, there are just two parameters for the underlying epidemiological model ( $\beta$ , the transmission rate;  $\mu$ , a measure of the infectious period). To these we added three parameters (Table 1) to allow for resistance. One or more of these could be set to one and effectively eliminated as a separate parameter. We also introduced a similarly tunable parameter ( $\rho$ ) for the ratio of resistant hosts in the population, while  $\lambda$  is a measure of uncertainty about whether or not disease is likely to occur. The other parameters listed in (Table 1), ( $R_0$ ) and ( $\xi$ ), are compound parameters derived from the others.

We have also used a strongly simplifying assumption that crop yield can be assessed from the final level or the integral over time of susceptible hosts, which assumes that infected hosts do not contribute to yield. Our intention here, however, was to focus on principles relating to decision-making in relation to the deployment of resistance. For this we have preferred to keep the model for the epidemic and for yield simple. It is possible that infected hosts might well contribute to yield and this could readily be included in the model as could additional feedback loops for the effects of different levels of infection on growth dynamics of the host [26, 32, 33].

Our results have shown the importance of accounting for uncertainty to inform policy on behalf of an agricultural planner. We restricted our analyses to considering the probability of an outbreak occurring. Future analyses could address the robustness of the conclusions to short-term fluctuations associated with demographic and environmental stochasticity. Of more likely importance, however, are longer-term fluctuations that result in periodic epidemics, arising from repeated introductions of the pathogen from outside the system, and periods of long-term environmental suitability. The intricacy of this approach lies in the fact that while the environmental suitability forces outbreaks with a certain period, introducing the resistant variety would change the intrinsic time-scale of the epidemic [19]. Combining these two effects could produce complex dynamics [19, 24]. Finally, it would be interesting to include the effects of heterogeneity in grower behaviour or a more formal treatment of risk-aversion [34, 35]. In this paper, we

model the yield in the presence of uncertainty as the average of the yield when the outbreak epidemic does or does not occur. However, farmers and policy makers tend to prefer control strategies that minimize the probability of losses rather than those that maximize the probability of gains [36, 37]. Our model could be combined with game theoretic approaches to analyse these outcomes.

## Supplementary Materials

### S1. Sketch explaining the yield approximation

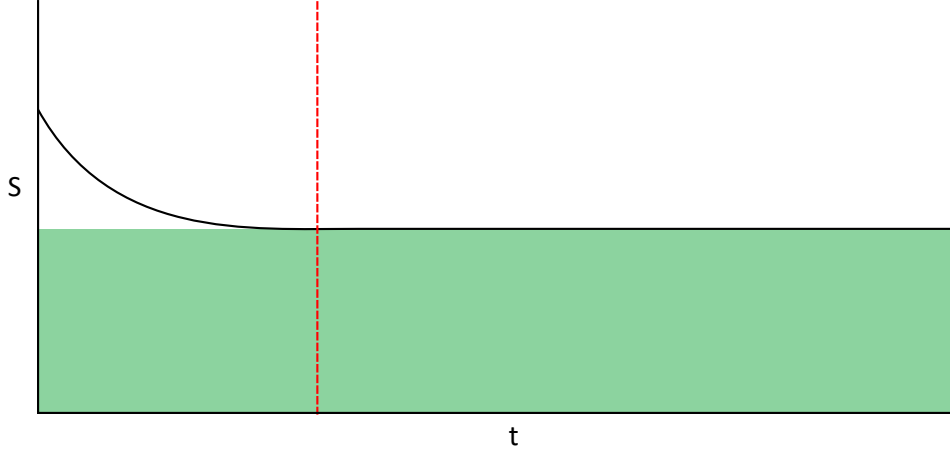


Figure 8: The yield is modelled as the area under the curve  $S(t)$ . In the figure, the vertical red line marks the end of the epidemic. When the yield is accumulated over a time period much longer than the duration of the epidemic, we approximate it by the area of the green rectangle, that is by the length of the time period multiplied by the surviving population,  $S(t = \infty)$ .

### S2. Calculation of the probability of a large scale outbreak

Let  $q$  be the probability of no outbreak after the first infection has occurred and furthermore, let  $q_R$  be this probability given the first infection was of a resistant host, and  $q_S$  the same for standard host. Then by the standard argument [38], assuming large populations sizes, we have

$$q = p_S q_S + p_R q_R \quad (17)$$

$$q_S = \frac{\mu}{\mu + \beta(1 - \rho + \eta\rho)} + \frac{\beta(1 - \rho + \eta\rho)}{\mu + \beta(1 - \rho + \eta\rho)} q_S q \quad (18)$$

$$q_R = \frac{\mu}{\mu + \beta\nu\sigma(1 - \rho + \eta\rho)} + \frac{\beta\nu\sigma(1 - \rho + \eta\rho)}{\mu + \beta\nu\sigma(1 - \rho + \eta\rho)} q_R q \quad (19)$$

where  $p_S$  and  $p_R$  are the probabilities of a newly infected host being of the standard variety or the resistant variety and are given by

$$p_S = \frac{1 - \rho}{1 - \rho + \eta\rho} \quad (20)$$

$$p_R = \frac{\eta\rho}{1 - \rho + \eta\rho}. \quad (21)$$

These are three equations for three unknowns  $q$ ,  $q_S$  and  $q_R$  and solving them leads to a quadratic equation for  $q$ . We have  $P_{\text{takeoff}} = 1 - q$  and this gives

$$P_{\text{takeoff}} = \frac{-1 - \frac{1}{\sigma\nu} + R_0 F + \sqrt{(\frac{1}{\sigma\nu} - 1)^2 + R_0^2 F^2 + 2R_0(\frac{1}{\sigma\nu} - 1)((F - 2\eta\rho))}}{2R_0 F} \quad (22)$$

where, for simplicity, we have written  $F = 1 - \rho + \eta\rho$ .

### S3. Calculation of the yield function from the final size equations

It can be readily checked that the system (1-5) has two conserved quantities  $X_1$  and  $X_2$  given by

$$X_i = -\ln S_i + R_0 \sum_j A_{ij}(I_j + S_j), \quad (23)$$

where  $A$  is a  $2 \times 2$  matrix

$$A = \begin{pmatrix} (1-\rho) & \nu\rho\sigma \\ \eta(1-\rho) & \nu\rho\eta\sigma \end{pmatrix}. \quad (24)$$

This gives the final size equations [38]

$$S_S^\infty = e^{-R_0(1-\rho)(1-S_S^\infty) - R_0\nu\sigma\rho(1-S_R^\infty)} \quad (25)$$

$$S_R^\infty = (S_S^\infty)^\eta. \quad (26)$$

Unsurprisingly,  $S_R^\infty > S_S^\infty$ . These can be used to eliminate  $S_R^\infty$ ; however, the resulting equation for  $S_S^\infty$  cannot be solved analytically. It is possible to make progress however. First note that the map between  $S_S^\infty$  and  $\rho$  is one-to-one as  $S_S^\infty(\rho)$  is strictly increasing. It is useful to introduce some additional notation. When  $\rho = 0$ , the final size  $S_S^\infty$  will be denoted by  $s_0$ , where

$$s_0 = e^{-R_0(1-s_0)}. \quad (27)$$

When  $\rho = 1$ , the final size  $S_S^\infty$  will be denoted by  $s_1$ , where

$$s_1 = e^{-R_0\nu\sigma(1-s_1^\eta)}. \quad (28)$$

Note that it is possible that  $s_1 < s_0$ , i.e. the proportion of the susceptible hosts in the standard population that survive the epidemic is lower with full control. This is not discussed in the main text because it has no impact on the general behaviour outlined in Figure 1. There are two cases to be considered. If  $R_0\nu\sigma \leq 1$ , the outbreak is prevented for  $\rho \geq \rho_c$  where

$$\rho_c = \frac{R_0 - 1}{R_0(1 - \nu\sigma\eta)}. \quad (29)$$

In this case the allowed range of values of the final size is  $s_0 \leq S_S^\infty \leq 1$ , corresponding to  $0 \leq \rho \leq \rho_c$ . On the other hand, if  $R_0\nu\sigma > 1$ , the outbreak cannot be prevented. The allowed range of values of the final size is then  $s_0 \leq S_S^\infty \leq s_1$ , corresponding to  $0 \leq \rho \leq 1$ . On these intervals the function  $S_S^\infty(\rho)$  can be inverted to give

$$\rho(S_S^\infty) = \frac{\ln(S_S^\infty) + R_0(1 - S_S^\infty)}{R_0[1 - S_S^\infty - \nu\sigma(1 - (S_S^\infty)^\eta)]} \quad (30)$$

and therefore the yield can be written as a function of  $S_S^\infty$ ,

$$Y(S_S^\infty; R_0, f, \nu, \eta) = S_S^\infty - (S_S^\infty - f(S_S^\infty)^\eta) \frac{\ln(S_S^\infty) + R_0(1 - S_S^\infty)}{R_0[1 - S_S^\infty - \nu\sigma(1 - (S_S^\infty)^\eta)]}. \quad (31)$$

### S4. Extreme control results

By scanning through possible values of the parameters, we have determined that the benefits of the intermediate control are not significantly higher than those of the extreme control. This is intuitive, since in Figure 2 we can see that the area where the intermediate control is optimal is small and so we would not expect the profit function to raise and drop too rapidly inside that area. Intuitively from Figure 2, the benefits of the intermediate control will approximately be the largest when  $Y_N = Y_F$ , where  $Y_N$  is the yield when we do not control at all and  $Y_F$  is the yield under full control. We have

$$Y_N = s_0 \quad (32)$$

$$Y_F = \begin{cases} 1 - \rho_{\text{full}} + f\rho_{\text{full}} & \text{if } R_0(1 - \xi) < 1 \\ fs_1^\eta & \text{if } R_0(1 - \xi) > 1 \end{cases} \quad (33)$$

where  $s_0$  and  $s_1$  are as above and  $\rho_{\text{full}} = \min(1, \rho_c)$ . Note that raising equation (28) to the power  $\eta$  reveals that  $s_1^\eta$  is a function of  $\xi$  and  $R_0$  only. Therefore  $Y_N = Y_F$  is represented by a curve in the  $(\xi, f)$  plane. Numerical analysis reveals that along this curve, the benefit of intermediate control is the greatest when  $\eta = 1$ . This is visually intuitive from Figure 2. Also, comparing the function  $Y_F$  and  $Y_N$  immediately gives the conditions (14) and (15), they both follow from rearranging  $Y_F > Y_N$ .

## S5. Numerical simulation: technical details

We ran 2000 simulations where the parameters  $\nu$ ,  $\eta$ ,  $\xi$  and  $f$  were selected uniformly at random from the interval  $(0.1, 1)$  and then discarded if they gave unreasonably high values of  $\sigma$  (above 20). To calculate the yield, the simulations were run with 10000 hosts, starting with 1 infected host (standard with probability  $1 - \rho$  and resistant with probability  $\rho$ ). Each simulation ended if either the outbreak has ended (the number of infected hosts dropped below 1) or if the duration of the epidemic reached the cut-off value of 1000 infectious periods. The cut-off value was introduced for the reasons of computational feasibility and since the duration of the outbreaks without any control was about 15 infectious periods, the cut-off does not have a significant effect on the analysis.

## Acknowledgements

We thank the four anonymous referees for their helpful comments and Matt Castle and Andrew Craig for many helpful suggestions, which we feel have greatly improved the paper.

## Authors' contributions

M.V. conceived and designed the study. M.V. and N.J.C. implemented the study. M.V. and C.A.G. wrote the initial version of the paper. All three authors discussed the results and their implications throughout the study, and edited the final version of the paper.

## Funding statement

M.V. was funded by Bakala Foundation and the Trinity College, Cambridge.

## References

- [1] Agrios G. Plant Pathology. 5th ed. Elsevier; 2005.
- [2] Oerke EC, Dehne HW, Schönbeck F, Weber A. Crop production and crop protection: Estimated losses in major food and cash crops. Elsevier; 1994.
- [3] Oerke EC. Crop losses to pests. The Journal of Agricultural Science. 2006 Feb;144(1):31–43.
- [4] Savary S, Ficke A, Aubertot JN, Hollier C. Crop losses due to diseases and their implications for global food production losses and food security. Food Security. 2012 Dec;4(4):519–537.
- [5] Madden LV. Botanical Epidemiology: Some Key Advances and its Continuing Role in Disease Management. European Journal of Plant Pathology. 2006 May;115(1):3–23.
- [6] Johnson R. A critical analysis of durable resistance. Annual review of phytopathology. 1984;22:309–330.
- [7] Mundt CC. Durable resistance: A key to sustainable management of pathogens and pests. Infection, Genetics and Evolution. 2014 Oct;27:446–455.
- [8] Jones JDG, Dangl JL. The plant immune system. Nature. 2006 Nov;444:323–329.

- [9] Crute IR, Holub EB, Burdon JJ, editors. The gene-for-gene relationship in plant-parasite interactions. CAB INTERNATIONAL; 1997.
- [10] Brown JKM. Durable resistance of crops to disease: a Darwinian perspective. *Annual Review of Phytopathology*. 2015;53:513–539.
- [11] Singh RP, Huerta-Espino J. Effect of Leaf Rust Resistance Gene Lr34 on Grain Yield and Agronomic Traits of Spring Wheat. *Crop Science*. 1997;37:390–395.
- [12] Burdon JJ, Barrett LG, Rebetzke G, Thrall PH. Guiding deployment of resistance in cereals using evolutionary principles. *Evolutionary Applications*. 2014 Jun;7(6):609–624.
- [13] Ellis JG, Lagudah ES, Spielmeier W, Dodds PN. The past, present and future of breeding rust resistant wheat. *Frontiers in Plant Science*. 2014 Nov;5(641).
- [14] Anon. AHDB Recommended lists for cereals and oilseeds. Agriculture and Horticulture Board, Stoneleigh; 2015.
- [15] Brown JKM. Yield penalties of disease resistance in crops. *Current Opinion in Plant Biology*. 2002 Aug;5(4):339–344.
- [16] Brown JKM, Rant JC. Fitness costs and trade-offs of disease resistance and their consequences for breeding arable crops. *Plant Pathology*. 2013 Dec;62:83–95.
- [17] Kjær B, Jensen HP, Jensen J, H JJ. Associations between three ml-o powdery mildew resistance genes and agronomic traits in barley. *Euphytica*. 1990 Apr;46(3):185–193.
- [18] van den Bosch F, Gilligan CA. Measures of Durability of Resistance. *Analytical and Theoretical Plant Pathology*. 2003 May;93(5):616–625.
- [19] Lo Iacono G, van den Bosch F, Gilligan CA. Durable Resistance to Crop Pathogens: An Epidemiological Framework to Predict Risk under Uncertainty. *PLOS Computational Biology*. 2013 Jan;9(1).
- [20] Papaix J, Goyeau H, Du Cheyron P, Monod H, Lannou C. Influence of cultivated landscape composition on variety resistance: an assessment based on wheat leaf rust epidemics. *New Phytologist*. 2011 May;191(4):1095–1017.
- [21] Fabre F, Rousseau E, Mailleret L, Moury B. Durable strategies to deploy plant resistance in agricultural landscapes. *New Phytologist*. 2012 Jan;193(4):1064–1075.
- [22] Forster GA, Gilligan CA. Optimizing the control of disease infestations at the landscape scale. *Proceedings of the National Academy of Sciences*. 2007 Mar;104(12):4984–4989.
- [23] Zadoks JC, Schein RD. *Epidemiology and plant disease management*. Oxford University Press Inc.; 1979.
- [24] Keeling MJ, Rohani P. *Modeling Infectious Diseases in Humans and Animals*. Princeton University Press; 2008.
- [25] Madden LV, Hughes G, van den Bosch F. *The Study of Plant Disease Epidemics*. The American Phytopathological Society; 2007.
- [26] Cunniffe NJ, Koskella B, Metcalf CJE, Parnell S, Gottwald TR, Gilligan CA. Thirteen challenges in modelling plant diseases. *Epidemics*. 2015 Mar;10:6–10.
- [27] Heffernan JM, Smith JR, Wahl LM. Perspectives on the basic reproductive ratio. *Journal of the Royal Society Interface*. 2005 Sep;2(4):281–293.
- [28] Hall RJ, Gubbins S, Gilligan CA. Evaluating the Performance of Chemical Control in the Presence of Resistant Pathogens. *Bulletin of Mathematical Biology*. 2007 Feb;69(2):525–537.



- [29] Parlevliet JE. Components of resistance that reduce the rate of epidemic development. *Annual Review of Phytopathology*. 1979 Sep;17(1):203–222.
- [30] Savary S, Bosc JP, Noirod M, Zadoks JC. Peanut rust in West Africa: a new component in a multiple pathosystem. *Plant disease*. 1988;72(12):1001–1009.
- [31] Park AW, Gubbins S, Gilligan CA. Invasion and persistence of plant parasites in a spatially structured host population. *Oikos*. 2003 Apr;94(1):162–174.
- [32] Bailey DJ, Gilligan CA. Modelling and analysis of disease-induced host growth in the epidemiology of take-all. *Phytopathology*. 2004 May;94(5):535–540.
- [33] Bailey DJ, Kleczkowski A, Gilligan CA. An Epidemiological Analysis of the Role of Disease-Induced Root Growth in the Differential Response of Two Cultivars of Winter Wheat to Infection by *Gaeumannomyces graminis* var. *tritici*. *Phytopathology*. 2006 May;96(5):510–516.
- [34] Gent DH, De Wolf E, Pethybridge SJ. Perceptions of Risk, Risk Aversion, and Barriers to Adoption of Decision Support Systems and Integrated Pest Management: An Introduction. *Phytopathology*. 2011 Jun;101(6):640–643.
- [35] Gilligan CA, Truscott JE, Stacey AJ. Impact of scale on the effectiveness of disease control strategies for epidemics with cryptic infection in a dynamical landscape: an example for a crop disease. *Journal of the Royal Society Interface*. 2007 Oct;4(16):925–934.
- [36] McRoberts N, Hall C, Madden LV, Hughes G. Perceptions of Disease Risk: From Social Construction of Subjective Judgments to Rational Decision Making. *Phytopathology*. 2011 Jun;101(6):654–665.
- [37] Cunniffe NJ, Stutt ROJH, DeSimone RE, Gottwald TR, A GC. Optimising and Communicating Options for the Control of Invasive Plant Disease When There Is Epidemiological Uncertainty. *PLoS Comp Bio*. 2015 Apr;11(4).
- [38] Diekmann O, Heesterbeek H, Britton T. Mathematical tools for understanding infectious disease dynamics. Princeton University Press; 2012.